图书情报工作
第64卷 第21期 2020年11月

德国科研数据基础设施的构建模式及启示

周雷[1,2] 杨萍[1,2] 袁汝兵[1,2]
[1] 北京市科学技术情报研究所 北京 100044 [2] 北京科技战略决策咨询中心 北京 100044

摘 要：［目的/意义］科研数据基础设施是科研数据高效管理和利用、数字化科研得以高效运转而不可或缺的支撑性基础设施。［方法/过程］以德国科研数据基础设施项目为研究对象，从国家地区任务内容区别、建设机构组织和相关管理机制角度对科研数据基础设施建设模式进行分析。［结果/结论］研究发现，德国科研数据基础设施在不同层面具有不同侧重，在国家层面以数据的标准、工具为主；在地方层面以服务、培训为主，在建设中采取广泛合作、能力互补的策略进行机构组织；在管理中，则具有侧重统一管理和监督的特点。最后对我国进行科研数据基础设施建设给予建议。

关键词：科研数据 基础设施 科研数据基础设施 构建模式
分类号：G259.1 G311
DOI：10.13266/j.issn.0252-3116.2020.21.017

1 引言

科研数据是科学技术研究、试验开发等过程中产生的数据以及通过观测监测、考察调查等方式取得并用于科学研究活动的原始及其衍生数据[1]。目前随着科学研究"第四范式"的到来，科研数据已成为信息化时代推动科技发展、社会进步的重要战略资源[2]，科研数据基础设施建设也就成为重要研究命题[3]。数据基础设施内容较为宽泛，按照维基百科的定义，任何一个可以促进数据分享与消费的电子化设施都是数据基础设施。因此，促进科研有效运转所需要的任何信息化服务与工具也都可以看作科研数据基础设施。由于目前科研数据的主要目的是分享与重用，所以，本研究所涉及的科研数据基础设施主要是指能够支持科研数据的分享、保存、获取、重用与验证的一系列支撑性信息基础设施。欧盟最早针对科研数据基础设施进行基础性研究，如2009年"科研数据高水平专家组(High-Level Expert Group，HLEG)"项目[4]、2010年"绘制欧洲科研基础设施蓝图"项目[5]、GRDI2020[6]、2011年"泛欧洲协作数据基础设施"项目[7]等，主要从科研数据基础设施建设的益处、困难以及行动路径进行论述。此后美国、澳大利亚等国家也逐步开始进行相关研究，并提出了众多科研数据基础设施框架模型[8]。但从实践中看，当前科研数据基础设施建设呈现出种类繁多、碎片化、变化不定的特点，组织管理方式也大相径庭。目前主要存在形式为单独构建和联盟构建两种方式[9]。单独构建模式最为常见，主要指科研机构将平时科研项目的数据单独整理归档。目前大部分机构采用这种方式共享科研数据，也是我国相关机构采用的主要模式。该模式一般数据量有限，很难保障持续性和完整性，数据提升价值空间不大。联盟构建主要是数据中心与高校/政府合作或者几个机构建立共享联盟，收集共享科研数据。如荷兰的4TU. Centre for Research Data，其是由荷兰4所高校图书馆共同建设的科研数据联盟[10]，美国高校校际政治与社会研究联盟(Inter-university Consortium for Political and Social Research，ICPSR)，由美国密歇根大学创建，有超过700家学术机构会员，超过50万份数据档案资料[11]。在我国，如清华大学中国经济社会数据研究中心[12]、中国人民大学中国调查与数据中心[13]也都属于这种构建模式。这种模式虽然在运营的连续性和数据的可靠性要高于第一种构建模式，但相对于海量的研究数据，其还是较为分散和孤立的。同时，在这种模式下，数据虽得到部分汇集，但也没有形成一套集数据使用、提供两终端和中间技术、法律为一体的科研数据增值服务流程。德国作为全球重要科研力量，是开放科学重要的发起国和参与国，从

作者简介：周雷(ORCID:0000-0003-2971-3138)，助理研究员，硕士；杨萍(ORCID:0000-0002-3827-3114)，主任，副研究员，博士，通讯作者，E-mail:yangp@bjstinfo.com.cn；袁汝兵(ORCID:0000-0002-0369-6950)，副所长，副研究员，硕士。

收稿日期：2020-04-20 修回日期：2020-07-24 本文起止页码：140-150 本文责任编辑：徐健

周雷，杨萍，袁汝兵.德国科研数据基础设施的构建模式及启示[J].图书情报工作,2020,64(21):140－150.

2013 年起,大量高校就开始开展科研数据管理服务,并建立起各自机构科研数据知识库,同时,一些以联盟形式构建的科研数据中心也逐步出现,如德国社会科学数据中心 GESIS[14],LMU-ifo 经济与商业数据中心[15]等。在此背景下,德国联邦和州两级政府开始构建联邦/州科研数据基础设施,并在项目进程中,逐步确定自身作用,形成一体的科研数据基础设施。本研究以联邦/州科研数据基础设施项目为研究对象,分析其构建机制,为我国科研数据管理及建立国家科研数据基础设施提供参考。

2 德国科研数据基础设施构建发展概述

德国科研数据基础设施早期也以机构科研数据知识库、专题数据知识库为主,如海德堡大学的 heiDATA[16]、慕尼黑大学的 Open Data LMU[17]、柏林洪堡大学多媒体知识库 Media Repository[18]等。此后,以机构联盟形式参加的科研数据知识库逐步建立,如德国社会科学数据中心 GESIS、LMU-ifo 经济与商业数据中心等。面对相对分散的信息和数据,德国联邦和州政府相继开展统一的科研数据基础设施发展项目,时间线如图 1 所示。2014 年 1 月德国第一个州级科研数据基础设施项目 bwFDM 项目由巴登－符腾堡州科学研究和艺术部承担[19]。该项目由州内 9 所高校联合执行,一期分为 bwFDM-Communities 和 bwFDM-Info 两个子项目,第一步主要涉及学科需求调研,第二步则针对这些需求,为科研人员建立完整的科研数据管理信息平台(forschungsdaten. info);二期 bw2FDM 提出建立州科研数据中心。2016 年 5 月,德国信息基础设施委员会(Rat für Informationsinfrastrukturen,RfII)提出建立国家科研数据基础设施(Nationale Forschungsdateninfrastruktur,NFDI)[20],并于 2017 年 4 月公布了第一个讨论稿文件。正式提出了 NFDI 建立的基础是科研数据相关方的广泛合作,并在现有基础上实现相互关联、补充[21]。2018 年 3 月该委员会发布第二个讨论文件《合作为契机》,明确了 NFDI 的设计目标和特征,并明确"NFDI 是面向数据生产者和用户的分布式的、联网的基础设施,通过可靠且可持续的服务满足德国科研数据管理的通用和特定主题需求[22]。"2018 年 8 月德国科学组织联盟发布《关于制定联邦－州建立国家研究数据基础设施协议的讨论文件》[23],并于同年 11 月由联邦/州两级政府主管机构组成的科学联席会议(Gemeinsame Wissenschaftskonferenz,GWK)批准通过。2018 年 12 月,RfII 发布第三次讨论稿[24],提出将利用联合体模式进行 NFDI 的构建,并于 12 月获得主管机构 GWK 的批准,并由德国研究理事会(Deutsche Forschungsgemeinschaft,DFG)组织实施[25]。在 NFDI 讨论进程中,各州也相继开展自己的科研数据基础设施建设(由于各州项目名称不一,笔者套用德语习惯用法,简称 LNFDI),配合 NFDI 推进整个德国的科研数据管理,如黑森州的 HeDFI 项目(Hessische Forschungsdateninfrastruktur"黑森州科研数据基础设施")(11 所州内高校参加)、北威州的 NFDI Landesinitiative 项目(通过 DH. NRW 组织在全州内高校/科研机构推行)等。目前德国在 re3data 上注册的数据知识库超过 400 个,涉及自然科学、人文社科、工程、生命科学,基本建立起了一个遍布全联邦的知识库系统;而在服务领域,自 2013 /2014 年比勒菲尔德大学、柏林洪堡大学、哥廷根大学、海德堡大学等第一批德国高校开始科研数据管理服务,也积累了较为丰富的实践经验,目前,超过一半的综合和理工大学提供科研数据管理服务;在技术层面,一些涉及工具软件(如数据管理计划软件 RDMO)、学科领域数据管理(如工程领域科研数据管理方案 DIPL-ING)等国家级项目陆续结束,具备了实施国家科研数据基础设施的基础。从以上分析可以看出,德国科研数据基础设施建设从构建维度看,可以分为国家层面和州(地方)层面建设;从构建途经看,基本以联盟/联合体的形式进行;而从构建手段看,基本以互联互通为主,补充为辅的方式拓展。

3 分析角度和数据来源

3.1 分析角度

科研数据基础设施构建的最终目的是实施,因此,笔者对科研数据基础设施构建模式研究的分析也立足于实施角度。目前较为成体系的科研数据基础设施框架包括欧盟委员会科研数据高水平专家组(High-Level Expert Group,HLEG)、澳大利亚科研数据基础设施委员会(Research Data Infrastructure Committee,RDIC)等机构制定的实施方案。欧盟委员会科研数据高水平专家组认为科研数据基础设施是一个流动的循环系统,其强调基础设施建设中的不同参与者以及数据与服务在其中应如何互相关联运行[26]。澳大利亚 RDIC 则更加强调效率,认为科研数据基础设施应当协调国家、领域以及机构等不同层次的合作,避免重复建设,减少浪费[3]。具体来看,数据提供方提供原始数据,并利用信息技术机构的技术将数据管理、整理、归集;数据使用方一方面提供明确的科研需求,一方面又利用这些数

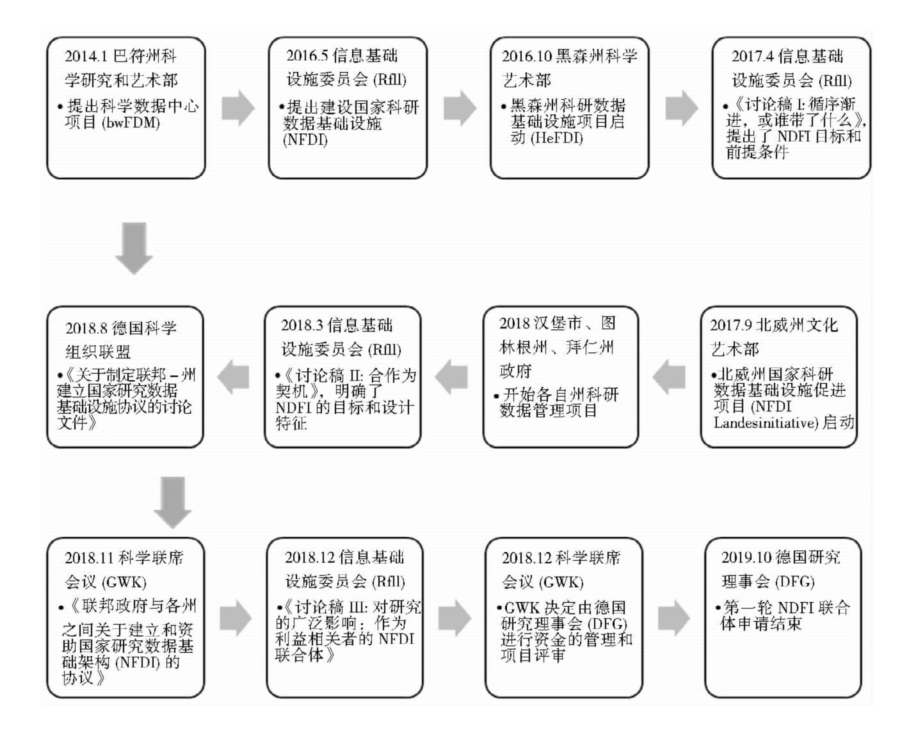


图 1　德国科研数据基础设施项目发展时间线

据展开研究获得成果，从而又可能成为数据的提供方；而信息技术机构则需要按照研究需要补充相关技术、软件、服务，既为现有研究提供便利，又成为其他机构参与科研数据生态机制中的有力工具。此外，在科研数据基础设施实施中，服务机构也尤为重要，一方面服务机构充当了数据提供方、数据使用方与信息技术、科研数据的中介，另一方面科研数据基础设施是一个开放系统，只有有了服务机构，才能吸引更多的科研人员、机构参与到数据的共享和基础建设之中。最后，现有数据知识库还需要通过研究机构和信息技术进行再次整合、联网，作为之前积累在生态系统内的数据资产被持续利用，如图 2 所示。而整个系统又要高效运作，避免资源的浪费。因此，科研数据基础设施构建的核心就是要明确设施中各元素的任务、承担任务的组织及相互间的关系，并且这个系统还要高效运转，减少无谓消耗。本研究也将立足于任务、承担机构以及保障高效运行的管理机制进行分析。

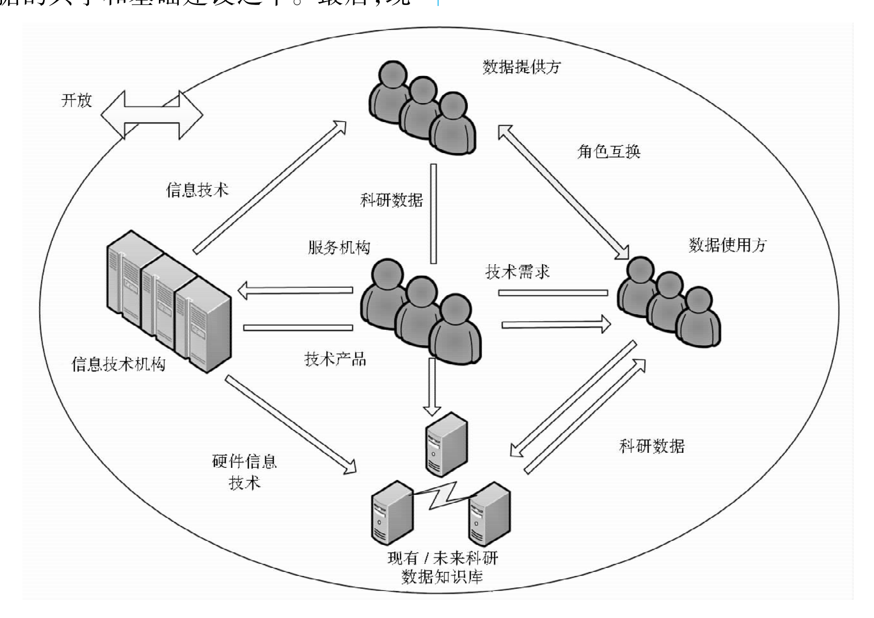


图 2　科研数据基础设施的实施方案

3.2 分析样本来源

德国科研数据基础设施项目已经全面开展,所以从项目的执行情况来分析其构建模式最为合适。目前,德国从联邦和州两级进行的科研数据基础设施建设项目,如表 1 所示。截至 2019 年底第一轮 NFDI 项目申请已经结束,共计 22 个[27];而在州层面,目前包括巴符州、黑森州等 7 个州已经开展科研数据管理基础设施项目,其中北威州、图林根州还主要以科研数据管理基础设施的先导类项目为主,而作为较早介入的巴符州和黑森州已经着手开展数据中心项目或相关基础设施的建设。

表 1　德国科研数据基础设施类项目列表

国家科研数据基础设施(NFDI)		州科研数据基础设施项目(LNFDI)
① 天文学、天体物理学和天体粒子物理学国家科研数据基础设施(Astro@ NFDI)	⑫ 个人健康国家科研数据基础设施(NFDI4Health)	① 巴符州 bwFDM 项目
② 光子和中子实验数据(DAPHNE)	⑬ 农业科学国家科研数据基础设施(NFDI4Agri)	② 黑森州 HeFDI 项目
③ 植物学研究基础数据(DataPLANT)	⑭ 医学信息学倡议和德国卫生研究中心(DZG)国家科研数据基础设施(NFDI4MED)	③ 北威州 Landesinitiative NFDI 项目
④ 德国人类基因组档案(GHGA)	⑮ 商业、经济及相关领域国家科研数据基础设施(BERD@ NFDI)	④ 图林根州 TKFDM 项目
⑤ 化学国家科研数据基础设施(NFDIChemie)	⑯ 催化技术国家科研数据基础设施(NFDI4Cat)	⑤ 汉堡市州 Open Science-Initiative 项目
⑥ 材料科学和技术及相关研究领域数据基础设施(FAIRmat-FAIR)	⑰ 地球系统研究国家科研数据基础设施(NFDI4Earth)	⑥ 拜仁州 fdm-Bayern 项目
⑦ 实验经济和社会学研究科研数据基础设施(ForumX)	⑱ 生物多样性、生态和环境国家科研数据基础设施(NFDI4BioDiversity)	⑦ 萨克森州 Saxfdm 项目
⑧ 社会、教育、行为和经济科学数据联盟(KonsortSWD)	⑲ 文化遗产国家科研数据基础设施(NFDI4Culture)	
⑨ 数学科研数据倡议(MaRDI)	⑳ 材料科学与工程国家科研数据基础设施(NFDI4MSE)	
⑩ 国家工程科研数据基础设施(NFDI4ING)	㉑ 粒子、天体粒子、强子与核物理加速技术国家科研数据基础设施(PAHN-PaN)	
⑪ 移动技术国家科研数据基础设施(NFDI4 MobilTech)	㉒ Text + 基于语言和文字的科研数据基础设施(Text +)	

4　国家/州科研数据基础设施的任务和相互关系

2018 年 8 月,德国科学组织联盟发布的《关于制定联邦－州建立国家研究数据基础设施协议的讨论文件》明确了德国科研数据基础设施建设将从国家和州(地方)两个层次配合实施。所以,分析首先应从最为宏观的国家和州各自定位和任务开始。笔者利用项目执行指导文件对国家/州两级科研数据基础设施的构建目的和作用进行分析。

4.1　NFDI——技术规范

按照项目文件,NFDI 的目的是系统化、标准化地管理科研数据,改变德国目前整个科学系统中相关数据分散且基于项目存储的情况,提供长期数据存储、备份和可访问,同时,在国内和国际上将数据联网,以确保第三方能够轻松、系统地访问这些数据,进行组合和分析[28],具体建设要求见表 2。按照该要求,NFDI 由多机构组成的项目联合体执行,其与学科内团体、系统外的科研数据管理服务机构及其他项目联合体密切合作条件下,开发包括数据采集、处理、标准、质量、管理服务、培训等一系列统一、通用的产品和服务(见图 3)。可以看出,NFDI 的构建核心是数据以及围绕这些数据制定的标准和必要技术元素的开发。首先,科研数据基础设施建设任务是保证数据在分布式存储知识库中可获取的基础上,实现数据在不同主体之间的流动。因此,建立数据采集、处理、交换标准是实现互联互通和数据共享的基础。同时,在科研数据基础设施中,科研人员是数据产生的来源,而数据又与论文文献不同,中间加工环节往往受专业因素的限制仍以科研人员为主,所以,针对数据产生的服务、培训也需要统一规范,从而保证数据的基础质量。因此,围绕专业或学科领域建立起一系列标准是国家层面需要解决的重要基础工作。

4.2　LNFDI——贴近服务

黑森州科研数据基础设施项目 HeFDI 开始于 2017 年,具有较为完整的研究成果,所以笔者以 HeFDI 项目为主要参考。HeFDI 的建设内容见表 3,其由州内 11 所高校组成的联盟执行。其中涉及数据知识库、管理计划软件工具等“偏硬”部分由联盟内 11 所高校合

表 2 NFDI 建设内容要求

NFDI 建设内容要求[28]	组织维度:目标/任务	技术维度:目标/任务(目的)
① 与学科内研究团体密切合作,建立数据处理标准、程序和准则	与学科内研究团体密切合作	建立数据处理标准、程序和准则
② 改善与德国学术研究系统之外的,并具有科研数据管理方面专业知识的合作伙伴之间的联系和协作	与德国学术研究系统之外的,并具有科研数据管理方面专业知识的合作伙伴之间的联系和协作	获取专业知识、经验和技术协助
③ 与其他项目联合体一起参与科研数据管理的开发,建立通用跨项目的服务和标准	与其他项目联合体合作	开发和建立通用跨项目的服务和标准
④ 制定跨学科的元数据标准	-	元数据标准的制定
⑤ 根据学科研究团体的需求,开发可靠且可互操作的数据管理措施和服务	学科研究团体需求	开发可靠且可互操作的数据管理措施和服务
⑥ 增加现有数据的可重用性,包括超越学科范围的应用	-	增加现有数据的可重用性

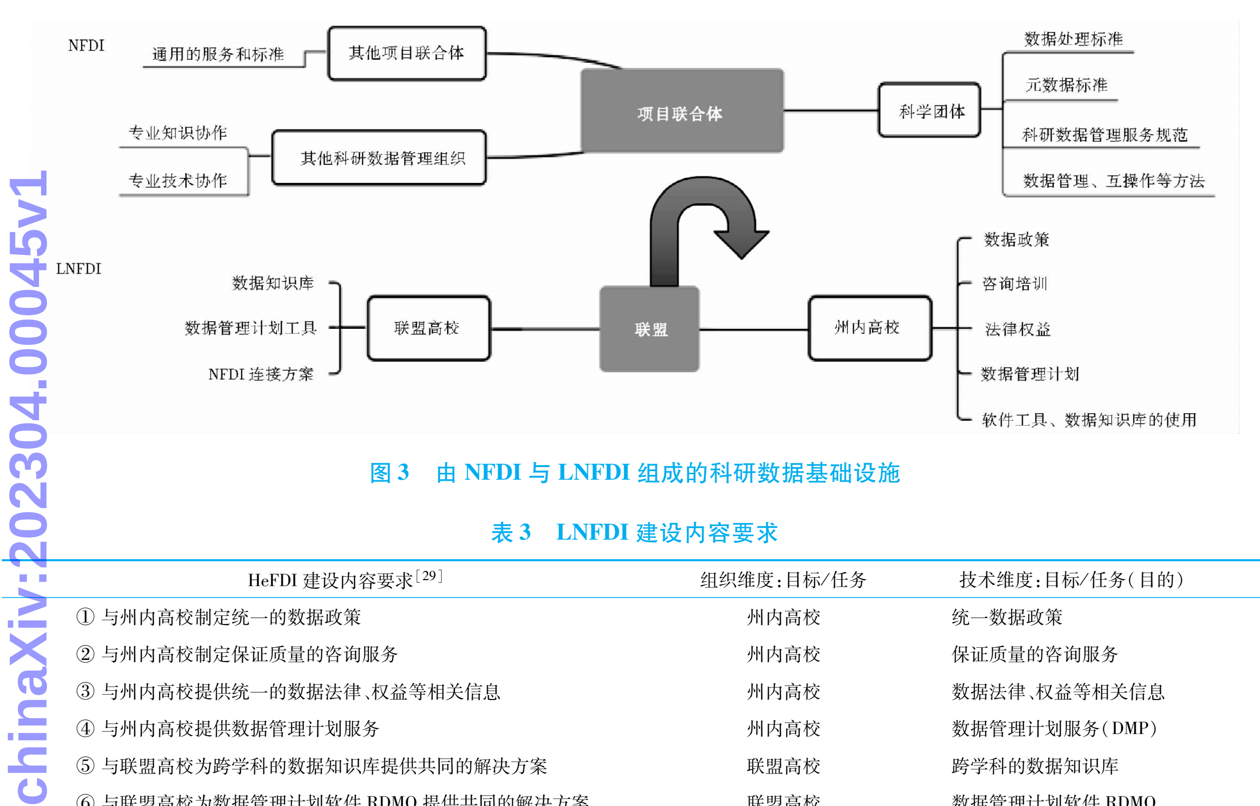


图 3 由 NFDI 与 LNFDI 组成的科研数据基础设施

表 3 LNFDI 建设内容要求

HeFDI 建设内容要求[29]	组织维度:目标/任务	技术维度:目标/任务(目的)
① 与州内高校制定统一的数据政策	州内高校	统一数据政策
② 与州内高校制定保证质量的咨询服务	州内高校	保证质量的咨询服务
③ 与州内高校提供统一的数据法律、权益等相关信息	州内高校	数据法律、权益等相关信息
④ 与州内高校提供数据管理计划服务	州内高校	数据管理计划服务(DMP)
⑤ 与联盟高校为跨学科的数据知识库提供共同的解决方案	联盟高校	跨学科的数据知识库
⑥ 与联盟高校为数据管理计划软件 RDMO 提供共同的解决方案	联盟高校	数据管理计划软件 RDMO
⑦ 与联盟高校共同开发与国家基础设施相关联的方案	联盟高校	与国家基础设施相关联的方案

作完成,而涉及科研数据政策、咨询服用等“偏软”部分则在州内高校统一协作下完成。HeFDI 项目着重于通过合作在州内开发统一的科研数据政策策略和咨询服务,以便在州内大学普及科研数据管理并提供相关服务。而在基础较好的 11 所高校则围绕统一的数据政策和模型建立包括数据管理计划软件、数据知识库等必要硬件设施,形成服务网络,结构见图 3“LNFDI”部分所示。因此,从建设形式上看,LNFDI 的构建方式与联邦层面一样,通过不同机构的相互组织合作,开发学科/主题统一的科研数据管理服务产品,如工具、政策、教学、咨询等。但与 NFDI 不同,LNFDI 的核心目的在于服务,并体现在服务的统一性、规范性上。目前,从现有科研数据管理服务上看,普遍以高校图书馆或者相关机构的信息中心为责任主体,自主开展如数据管理计划、咨询服务、最佳实践原则等具体服务。所以对于这些基层服务,LNFDI 的建立可以直接解决可能存在的服务规范不统一或是由于机构对于科研数据参与程度的差别造成的服务深度不一的问题,从根本上保障科研数据的质量。

4.3 NFDI 与 LNFDI 的关系及特点

4.3.1 相互辅助

如上文分析,NFDI 的核心是数据。其主要解决数据的共享和使用,建立层面是宏观的;LNFDI 和核心是服务,其主要解决如何为科研人员建立统一的服务模式和产品,建立层面是中观/微观的。这两方面是相互辅助相互支撑的。首先,NFDI 提供数据和技术的保障。NFDI 为学科提供广泛的基础研究数据、元数据标

周雷，杨萍，袁汝兵.德国科研数据基础设施的构建模式及启示[J].图书情报工作,2020,64(21):140－150.

准和一系列规范的技术服务元素，同时与系统外的科研数据管理技术机构合作，为国内在科研数据的使用上打通了如标识、版权的通道；其次，LNFDI 为 NFDI 提供基础。科研数据基础设施的根本是数据，除了现有的大装置、大项目等基础数据外，大量的科研项目数据是科研数据的主体，也是数据汇交价值提升的重要来源。而其基础都在于数据的共享和规范，而 LNFDI 建有众多机构层级的科研数据知识库，并且其建设主体一般在高校，所以，LNFDI 又为 NFDI 提供原始的支撑，为 NFDI 增添了数据来源的渠道，使 NFDI 不会从"零"起步；再次，LNFDI 为 NFDI 提供试验原型。数据的汇交、跨学科的应用，来源于科研项目的实践，而高校等基层科研机构可以使其需求更加容易辨识，同时相应数据量规模也便于开展相关技术的研发；最后，LNFDI 为 NFDI 提高了科研人员的数据文化和素养。完善的科研数据基础设施需要科研人员不断分享自己的科研数据，而 LNFDI 的服务既增强了科研人员的数据素养，又为科研人员参与数据共享提供了便利的条件，为 NFDI 的持续建设提供了源动力。

4.3.2 形成国家－地区两级科研数据基础设施生态体系

综上分析可以看出，德国实现了国家－地区两级科研数据基础设施生态体系的构建。在硬件层面，NFDI 提供国家基础设施，满足了高数据量、科研基础性数据的存储保障，LNFDI 提供地区层面的基础设施，其更加贴近基层科研人员，满足日常数据的共享，两者又通过连接方案实现互通；而在软件层面，NFDI 提供包含元数据标准、处理流程、培训咨询等学科领域标准、科研数据产品和服务，为整个生态系统提供了信息技术，保障数据的有效产生，实现数据的交换流动，而 LNFDI 可以使用这些成果提供相关服务，既可以在微观层面服务科研人员和机构，又可以吸纳越来越多的人员和机构共享科研数据，不断完善、扩大整个生态系统，加速数据生产者和使用者的转换，如图 3 所示。

5 组织构成

组织是为了实现共同目标，按照一定的结构形式、活动规律结合起来的系统，是科研数据基础设施的实施主体。在明确各项任务之后，组织办法是具体工作任务如何进行分工、分组和协调合作的基础。因此，通过分析其具体机构的组织办法、组织机构，可以了解系统内部之间的关系，以及组织的运转模式。笔者以 NFDI4Culture[30] 项目（文化遗产国家科研数据基础设施）为例进行分析，详细明确各组织及相关管理机制。

5.1 NFDI 的组织构成——广泛合作

从项目文件可知，NFDI 执行层面的主体是项目联合体，是承担建设科研数据基础设施的责任主体。而广泛合作既是项目要求，也是未来科研数据基础设施能够应用的基础。NFDI 的执行联合体可以分为 3 部分：主执行方、联合执行方和参与方。其中前两者作为项目执行机构参与基础设施的开发，而参与方作为相关辅助，在开发过程中起到补充特定学科、信息基础设施上专业知识和产品试用的作用。以 NFDI4Culture 项目联合体为例（见图 4），该项目联合体将建设研究主题为：建筑学、艺术史、音乐学、戏剧、电影和媒体科学领域内的科研数据基础设施。主申请人为德国美因茨科学和文学院。联合申请人中，有 4 所大学、3 家信息技术研究机构，1 家综合性文化机构，参与人有 14 个相关学会团体、26 个专业研究所和 25 家其他国际数据管理相关组织，如 UNESCO、Getty、PHAROS、ORCID、DataCite、Wikimedia 等。联合体按照各个机构的优势分配不同的任务，如柏林普鲁士文化遗产基金会是德国最大的文化机构，为整个生态系统提供了丰富的基础数据资源；德国国家科学技术图书馆（Technische Informationsbibliothek，TIB）等信息机构则主要承担基础科研数据的数字化任务。

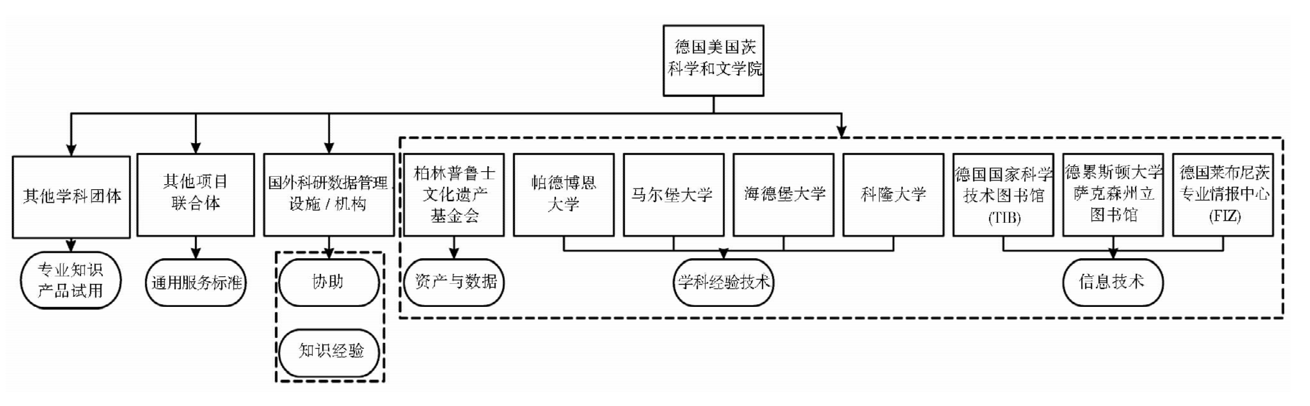


图 4　NFDI 组织结构

与学术研究团体、其他数据管理组织和其他项目联合体合作是项目的建设要求。在具体工作中,特点主要表现在:①与学科内研究团体的合作全面。除了联合申请人中的参与执行的4所高校外,参加人中有14个学会团体、26家学科研究机构,涵盖了德国整个相关学科体系。由于研究团体是科研数据的最终使用者和生产者,所以,项目成果的应用具有更加广泛的适用性,也有利于产品的推广和普及;②国际科研数据管理组织必不可少。在参加人中,有25个国际数据管理相关组织,因为从元数据方案、数据的出版和标识等众多领域都有一定的基础和规定,如最为常用的数字对象唯一标识符DOI、开放学者身份标识ORCID,或是一些特点学科的标识系统等。此外,一站式的发现系统也是任何资源设施的必要组成,因此,和国外一些数据仓储的合作也是必须的。所以,从技术上,国际上具有数据管理方面专业知识的合作伙伴可以为联合体提供相应的协助和经验,而从应用上,也为增强科研数据影响力、提高科研人员参与度提供前提条件;③与其他项目联合体间注重通用信息技术的合作。在NFDI4Culture项目合作意向书中[31],明确NFDI4Culture中的文化遗产数据与MaRDI(数学科研数据计划)、NFDINeuroscience(国家神经科学研究数据基础架构)、Text+(基于语言和文字的科研数据基础设施)分别在数学建模、图像、文字的编码和元数据方面具有共性技术,因此,对通用技术的合作是满足开发和建立统一服务和标准的前提要求。其实,在科研数据基础设施中,不仅在编码、元数据、标识符、互操作、专用术语等方面,在硬件系统(如统一身份访问管理平台、云基础设施、计算能力)、法律和版权(如敏感数据、商用数据),以及数据的治理领域(如数据文化、奖励机制、出版资助等)都存在广泛的共性技术。因此,和其他项目联合体合作是有效降低成本的重要手段。

5.2 LNFDI的组织构成——以地区图书馆/计算中心为中心

由于LNFDI的核心是服务,目标是协同统一,所以与NFDI不同,在LNFDI中,各个高校的图书馆和计算中心是项目的主要承担者,组织的设计就是如何保障不同高校相关部门的协同合作。具体来看,HeFDI项目工作组由地区(州)内11所高校的图书馆、计算中心以及学科专业代表构成,协调州内各高校的科研数据管理服务,其主要作用是保障技术标准的统一和服务策略的规范。此外,州图书馆信息技术服务联盟HeBIS为工作组提供技术支持,而项目各参与高校主管副校长共同组成的指导委员会协调各高校/机构的顶层科研数据政策和负责项目进度联络工作,保障组织运行的框架一致。

6 高效保障机制

整体科研数据基础设施系统需要高效运行的保障机制。首先,由于NFDI采用项目基金申请的方式进行,所以,项目管理体制也是项目有效运行的保证;其次,在生态系统中需要大量数据知识库,其建设过程耗费大量人力物力,且又不存在国家层面的统一规划,因此,数据知识库的使用机制也是减少资源浪费的有效手段。

6.1 管理机构——统一管理、加强协调和监管

NFDI的治理采用战略、执行、监管三角结构,具体包括项目联合体大会、学科理事会、董事会和专家委员会[32](见图5)。其主要特点包括:①在战略层面,学科理事会保证项目的先进性和实用性。学科理事会是战略决策机构,其由联合体大会、德国科学院联盟、科学联席会议(GWK)及董事会组成。其中,德国科学院联盟由10家全国性科研机构、学术组织所组成,集合了德国几乎所有研究机构,是科学需求/利益的代表,同时,由于科研数据的共享,又涉及学术承认、奖励刺激机制,所以,NFDI也需要得到科学届的支持。GWK是政府组织、资助机构,是出资方的代表,也是项目利用价值的关切方,起到了财政资金的监管作用。而这两者在业务上由董事会提供支持。董事会由专业信息机构组成,具有专业的资源服务、信息技术能力,因此可以增强策略的可执行性,为NFDI的战略目标做出规划,既体现科研需求,又符合国家战略方向。②在执行层面,注重项目建设协调,减少重复建设,保证治理统一。项目联合体大会由各个联合体派出代表组成,其主要目的是对各个学科科研数据基础设施开发进行内容和技术上的协调与统一,正如上文所述,科研数据基础设施除技术外,在治理、法律等方面都有诸多共同的框架性问题,所以联合体大会不仅加强了服务标准规范性,避免重复建设,还可以保证数据治理的统一性。③在监管层,设立独立的监管机构。德国研究理事会(DFG)任命专家委员会,进行资助申请评估,而项目执行的监督则由DFG负责,其与具体管理机构并不存在直接的隶属关系,保证了监管相对独立,效力客观可靠。而德国研究理事会(DFG)也是德国最大的科研资助机构,对项目资助具有丰富的经验。

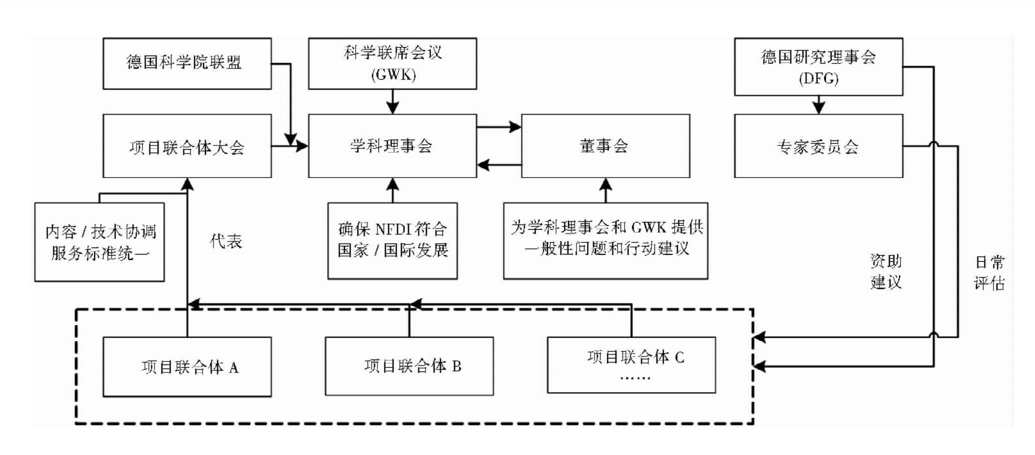


图5　NFDI管理机制

6.2　数据知识库建设——公用模式

构建区域内基础数据知识库，既是其中心任务要求之一，也是与NFDI互联互通的重要节点。笔者以HeFDI在数据知识库使用机制为分析对象，提供借鉴参考。HeFDI主要采用机构数据知识库合作公用的模式满足不同机构的使用需求，即主要由州内部分有条件的高校进行数据知识库建设，其他合作高校参与使用，并按照使用量支付建设费用。采用这种模式好处和必要性在于：①降低了使用专用数据知识库的门槛。部分科研数据产生量较小的高校不必自己单独建设机构数据知识库。②减少了数据知识库互联的难度。避免了大量小型知识库，也使得每个知识库的质量有所保障。③发挥现有数据知识库的作用。有些高校已经建有数据知识库，并且形成了专业的服务团队和模式，所以可以得到充分利用。

7　借鉴和启示

目前，从政策层面看，我国已形成由以政府、行业机构和领域数据中心为主体的数据政策体系；从共享实践层面看，既包括政府、科研机构，也囊括企业社会力量。但正是由于这种相伴而生的共享实践，导致我国科研数据基础设施整体规划有所不足。同时，大部分科学数据依然保存在项目组或个人手中，即便相关机构建有科研数据中心等基础设施，也存在诸如重复建设、数据量有限、重建轻用等问题。而有些科学数据虽然进行集中管理，但由于技术、标准不一，各组成依然彼此相互孤立，数据兼容困难，价值提升有限。所以，德国科研数据基础设施的构建模式对我国相关建设具有较强的借鉴意义。

7.1　科研数据基础设施建设亟待体系性建设

目前，我国科研数据基础设施主要来自个别高校、中科院系统以及部分专业科研院所。其中，最为重要的组成是依托国家科技资源共享服务平台，形成的包括“国家高能物理科学数据中心”等20个国家科学数据中心和30个资源库[33]。此外，以武汉大学、复旦大学为代表的高校也逐步建立起数据管理平台，并协作推进科研数据管理[34]。但无论是国家科学数据中心，还是其他机构建立的数据知识库，从学科构成看，主要以基础学科为主，工程类、社科类数据偏少；从数据来源类型看，也侧重于基础性数据，前沿科学数据较少。反观德国科研数据基础设施建设，从横向——学科角度看，既包括基础性数据，如物理、生物、生态，又包括技术类数据，如材料、工程、医药健康，还有诸如社会科学、文化遗产等社会文化数据，体系上更加完整；从纵向——建设层级看，既有国家科研数据基础设施，又有地方科研基础设施，并通过数据知识库进行关联，构建网络，既避免的了重复建设，又可以有效进行数据互操作。2018年3月，国务院办公厅印发《科学数据管理办法》（以下简称《办法》）[35]，明确了科研数据战略资源地位，同时提供了治理依据。2020年4月，《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》发布[36]，明确了作为数据要素重要组成之一——科研数据的经济社会资源价值，所以，参考德国模式，建立国家科研数据资源目录，规范数据知识库学科专业体系，在具体学科专业内采用国家数据总站-地区/机构数据分站/结点的模式，全面加快我国科研数据基础设施建设。

7.2　加强科研数据基础设施的生态系统建设

数据的广泛是科研数据基础设施的基本保障。同时科研数据的共享不可能一蹴而就，随着开放科学影响力的加深，一定会有更多的学者、机构参与到科研数据共享中来。因此，科研数据基础设施一定要保持开放性，要保持开放性就要有一定的流程、标准、软件和工具来形成开放的接口。同时，人是科研数据的来源，

所以,统一、规范的咨询服务、科研数据管理的素养教育也是必要的组成元素,形成科研数据文化。因此,进行科研数据基础设施建设其本质是建立起集数据、服务、标准、咨询、培训的一整套产品群,继而形成开放式的科研数据生态系统。目前,随着我国相关机构建立起科研数据平台,一批围绕科研数据的标准/规范也相继发布,如国家科学数据中心及各平台的学科核心元数据标准、数据模式描述规则和方法,中科院科研数据标准体系等[37],在技术层面已经有了一定基础,但技术的推广、优化在于频繁的使用。所以,参考德国模式,首先,将咨询、培训、服务也纳入科研数据中心建设的框架内,再从咨询、培训的内容、方法、资质层面进行开发,进而在应用层面,资助、鼓励机构图书馆/计算中心进行试用推广,普及科研人员/学生的科研数据管理素养。

7.3 建设需区别对待国家和地方建设任务和方法

科研数据基础设施的核心是数据,而数据又有不同来源,既有大尺度高数据量且相对静态的科研基础类数据,又有以科研项目等为代表的研究前沿小规模且相对动态的数据。同时,围绕这些数据又衍生一系列工具、服务、培训产品,而这些产品又面对不同对象,所以在科研数据基础设施建设层面需要区别对待国家和地方、专业和一般的建设任务和方法。在德国科研数据基础设施建设中,国家层面主要面对基础数据的建设,并开发出相关元数据标准,打通一体化服务模式等;而在地方层面,数据建设更加面向科研项目,并通过协调统一的科研数据管理计划、咨询服务等完善微观层面科研数据管理的服务和普及,为国家科研数据基础设施提供数据层面的补充,保障科研数据的来源和质量。《办法》指出,我国科研数据管理工作实行国家统筹、各部门与各地区分工负责的体制。所以,在国家层面,应推进建立不同学科或相近学科方向的国家科研数据中心,并着力打造相关的数据标准、工具、软件。在地方层面,优先资助地方机构图书馆/计算中心组成联盟,加强统一服务;对于部分有科研数据基础设施平台或有条件的高校或科研机构,鼓励其承担或成立地方层面(如省级)科研数据中心,而其他单位则通过服务购买的方式,补足资源消耗费用。这样既可以将数据有效集中,又可以避免重复建设,对于与国家平台互操作,也极为便利。

7.4 加强科研数据基础设施建设的一体化发展

目前,我国科研数据基础设施的构建主要为机构自建或国家统筹。以国家科学数据中心为例,其构建机制主要为科技部、财政部承担宏观管理,国家科技基础条件平台中心承担共享网的建设、运行、考核、评价等管理工作。而对于各个数据中心,主要是以中科院、国家部委为主管部门,下属专业院所为责任主体进行建设。利用这种模式可以快速有效建立基础性科研数据。但科学数据的应用价值在于互联互通,而我国科研数据基础设施的建设实体(如国家、地方或高校)相对分散,所使用的标准规范也不统一,数据价值提升有限。所以,协同发展全国科研数据基础设施建设至关重要。参考德国模式,在管理上,组建战略管理机构,由学科、信息服务、资源代表组成,既可以统一协调各数据中心,避免重复建设,又可以增强合作,对于共性技术给出统一的解决方案,增强技术的一般性;在建设上,采用项目资助形式,设立专项资金,既便于让有资质、有意愿的科研单位联合申请,又有利于社会上有资源、有技术的机构参与,形成国家/地方科研数据基础设施建设主体。参考基金管理办法,各级科技管理部门为资助责任主体,基金委或基金管理机构为监管主体,建设前加强评估,建设中加强监管,建设后加强考核,减少可能的重建轻用,大量"静态"平台,以及数据更新较少的状况,而条件不成熟的也可以暂缓建设。

7.5 加强科研数据基础设施建设的合作范围

科研数据基础设施建设需要广泛的合作,这不仅是技术能力的要求,也是未来数据中心、产品增强应用普及的有效手段。目前,我国科研数据中心基本以部门系统为主线、以数据为核心进行构建,相关成果在应用中的广泛性和适用性不足,数据也相对孤立,发现系统不足。而国外已经形成诸如 Dryad、Genbank 等在专业领域内的权威数据知识库,并与期刊、学者、研究机构广泛合作。参考德国模式,首先,在建设中增强和研究机构团体的广泛合作,一方面可以推广自身产品,加强应用,另一方面也可以加速其从使用者到贡献者的转化;其次,加强通用信息技术的合作,如德国就规定,必须识别出不同联合体间可能存在的通用技术,一方面可以降低开发难度,另一方面也便于数据的互联检索。此外,在国家平台建设上,应着力增强与国外科研数据基础设施管理组织的合作,提升管理运营水平,增强国际性和权威性。首先,国外在科研数据基础设施建设中,特别是数据中心建设、数据标准领域上已经积累了丰富经验,其次,一些涉及数据知识库、数据出版领域的认证、服务在国际上都有较为权威的体系;最后,加强数据关联、一站式检索也需要不同国家间不同机构的合作。在地方层面,加强服务机构的合作,鼓励

利用现有联盟，如高校图书馆联盟，开展统一的科研数据管理服务。

参考文献：

[1] 蔚海燕，范心怡. 研究型图书馆数据馆员能力需求及服务内容研究[J]. 图书馆，2019(4):85－91.

[2] 刘兹恒，曾丽莹. 我国高校科研数据管理与共享平台调研与比较分析[J]. 情报资料工作，2017(6):90－95.

[3] 王敬，王彦兵. 国外科研数据基础设施研究及实践的调研与分析[J]. 情报资料工作，2016(6):99－104.

[4] GRDI2020. Unlocking the full value of scientific data[EB/OL]. [2020－01－20]. http://www.grdi2020.eu/Pages/Unlock.aspx.

[5] European Science Foundation. MERIL Portal[EB/OL]. [2020－01－23]. http://portal.meril.eu/converisesf/static/about.

[6] FAAS H. GRDI2020[EB/OL]. [2020－01－05]. http://www.grdi2020.eu/Default.aspx.

[7] EUDAT Ltd. EUDAT Collaborative Data Infrastructure[EB/OL]. [2020－01－20]. http://eudat.eu/.

[8] BMBF. Nationale Forschungsdateninfrastruktur[EB/OL]. [2020－01－04]. https://www.bildung-forschung.digital/de/nationale-forschungsdateninfrastruktur-2542.html.

[9] 储文静，李书宁. 我国科学数据联盟管理模式构建研究[J]. 图书馆学研究，2019(14):51－57.

[10] TU Delft Library. 4TU. Centre for Research Data[EB/OL]. [2020－01－05]. https://researchdata.4tu.nl/en/.

[11] University of Michigan. ICPSR[EB/OL]. [2020－03－31]. https://www.icpsr.umich.edu/icpsrweb/.

[12] 清华大学. 清华大学中国经济社会数据研究中心[EB/OL]. [2020－01－05]. http://www.tcdc.sem.tsinghua.edu.cn/.

[13] 中国人民大学. 中国人民大学中国调查与数据中心[EB/OL]. [2020－01－05]. http://nsrc.ruc.edu.cn/.

[14] Leibniz-Institut für Sozialwissenschaften. GESIS[EB/OL]. [2020－01－20]. https://www.gesis.org/home.

[15] Ifo Institute. LMU-ifo Economics & Business Data Center[EB/OL]. [2020－01－20]. https://www.ifo.de/en/EBDC.

[16] Heidelberg University. heiDATA[EB/OL]. [2020－01－20]. https://heidata.uni-heidelberg.de/.

[17] Ludwig-Maximilians-Universität München. Open Data LMU[EB/OL]. [2020－01－20]. https://data.ub.uni-muenchen.de/.

[18] Humboldt-Universität zu Berlin. Media Repository Humboldt-Universität zu Berlin[EB/OL]. [2020－01－20]. https://media.hu-berlin.de/.

[19] Karlsruher Institut für Technologie. bwFDM[EB/OL]. [2020－01－20]. https://bwfdm.scc.kit.edu/.

[20] Rat für Informationsinfrastrukturen. Leistung aus Vielfalt[EB/OL]. [2020－01－20]. http://www.rfii.de/? wpdmdl=1998.

[21] Rat für Informationsinfrastrukturen. RFII-DISKUSSIONSPAPIER "SCHRITT FUR SCHRITT - ODER: WAS BRINGT WER MIT?" [EB/OL]. [2020－01－20]. http://www.rfii.de/download/rfii-diskussionspapier-2017/.

[22] Rat für Informationsinfrastrukturen. RFII-DISKUSSIONSPAPIER "ZUSAMMENARBEIT ALS CHANCE"[EB/OL]. [2020－01－20]. http://www.rfii.de/download/rfii-diskussionspapier-maerz-2018/.

[23] Allianz der Wissenschaftsorganisationen. Diskussionspapier zur ausarbeitung einer bund-Länder-vereinbarung zum aufbau einer nationalen forschungsdateninfrastruktur[EB/OL]. [2020－01－20]. https://www.fraunhofer.de/content/dam/zv/de/ueber-fraunhofer/wissenschaftspolitik/10/20180808% 20Diskussionspapier% 20NFDI% 20der% 20Allianz.pdf.

[24] Rat für Informationsinfrastrukturen. Wide impact for research: NFDI Consortia as Stakeholders[EB/OL]. [2020－01－20]. http://rfii.de/? wpdmdl=3818.

[25] DFG. Agreement between the federal government and the Länder concerning the establishment and funding of a National Research Data Infrastructure (NFDI)[EB/OL]. [2020－01－20]. https://www.dfg.de/download/pdf/foerderung/programme/nfdi/blv_en.pdf.

[26] Research data e-infrastructures: framework for action in H2020[EB/OL]. [2020－03－24]. https://indico.cern.ch/event/233119/contributions/492496/attachments/388387/540129/H2020-Framework-4-Action.pdf.

[27] DFG. NFDI statistik antragseingang[EB/OL]. [2020－01－20]. https://www.dfg.de/download/pdf/foerderung/programme/nfdi/191212_nfdi_statistik_antragseingang.pdf.

[28] DFG. Guidelines for Consortia National Research Data Infrastructure (NFDI)[EB/OL]. [2020－10－12]. http://www.dfg.de/formulare/nfdi100/nfdi100_en.pdf.

[29] Philipps-Universität Marburg. HeFDI[EB/OL]. [2020－01－20]. https://www.uni-marburg.de/de/forschung/kontakt/eresearch/projekte-und-netzwerke/hefdi.

[30] NFDI4Culture. NFDI4Culture[EB/OL]. [2020－01－15]. https://nfdi4culture.de/.

[31] DFG. NFDI4Culture - Consortium for research data on material and immaterial cultural heritage[EB/OL]. [2020－01－20]. https://www.dfg.de/download/pdf/foerderung/programme/nfdi/absichtserklaerungen/2019/2019_nfdi_4culture.pdf.

[32] DFG. Nationale Forschungsdateninfrastruktur[EB/OL]. [2020－01－07]. https://www.dfg.de/foerderung/programme/nfdi/index.html.

[33] 科技部. 关于发布国家科技资源共享服务平台优化调整名单的通知[EB/OL]. [2020－01－05]. http://www.most.gov.cn/mostinfo/xinxifenlei/fgzc/gfxwj/gfxwj2019/201906/t20190610_147031.htm.

[34] 屈宝强. 中国科学数据基础设施建设及发展对策研究[J]. 情报工程，2020，6(1):11－21.

[35] 国务院办公厅. 科学数据管理办法[EB/OL]. [2020－04－

02]. http://www. gov. cn/zhengce/content/2018-04/02/content_5279272. htm.

[36] 中共中央国务院. 中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见[EB/OL]. [2020 - 04 - 09]. http://www. gov. cn/zhengce/2020-04/09/content_5500622. htm.

[37] 胡良霖,郑晓欢,朱艳华,等. 中国科学院科学数据标准体系研究与实践[J]. 图书馆,2019(11):6 - 10.

作者贡献说明:

周雷:论文选题、资料收集与分析,论文撰写;

杨萍:论文研究框架指导、论文修改定稿;

袁汝兵:参与论文修订。

Enlightenment and Construction Mode of German Research Data Infrastructure

Zhou Lei[1,2] Yang Ping[1,2] Yuan Rubing[1,2]

[1] Beijing Institute of Science and Technology Information, Beijing 100044

[2] Beijing Science and Technology Strategic Decision Consulting Center, Beijing 100044

Abstract: [**Purpose/significance**] The research data infrastructure is an indispensable supporting infrastructure for the efficient management, utilization, and efficient operation of digital scientific research. [**Method/process**] This paper took the German research data infrastructure projects as the research object, and analyzed the construction mode from the perspectives of national and regional task content, construction organisation and related management mechanisms. [**Result/conclusion**] The study found that the German research data infrastructure has different emphasis at different levels. At the national level, data standards and tools are the main focus, while at the local level, services and training are the main focus. It adopts a strategy of extensive cooperation and complementary capabilities in construction. And in management, it has the characteristics of focusing on unified management and supervision. This paper concludes with suggestions for China's research data infrastructure construction.

Keywords: research data　infrastructure　research data infrastructure　construction mode